Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2-5 November 2003

DIMENSIONALITY REDUCTION FOR DENIAL OF SERVICE DETECTION PROBLEMS USING RBFNN OUTPUT SENSITIVITY

WING W. Y. NG, ROCKY K. C. CHANG, DANIEL S. YEUNG

Department of Computing, The Hong Kong Polytechnic University, Hong Kong
E-MAIL:{cswyng, csrchang, csdaniel}@comp.polyu.edu.hk

Abstract:

In this paper, we have presented a feature importance ranking methodology based on the stochastic radial basis function neural network output sensitivity measure and have shown, for the 10% training set of the DARPA network intrusion detection data set prepared by MIT Lincoln Labs, that 33 out of 41 features (more than 80% dimensionality reduction) can be removed without causing great harm to the classification accuracy of denial of service (DoS) attacks and normal packets (false positives rise from 0.7% to 0.93%). The reduced feature subset leads to more generalized and less complex model for classifying DoS and Normal. Exploratory discussions on the relevancy of the selected features and the DoS attack types are presented.

Keywords:

Denial of Service (DoS); Network Intrusion Detection; Feature Selection; Sensitivity Analysis; RBFNN

1 Introduction

According to the US based Computer Security Institute/Federal Bureau of Investigation's Computer Crime and Security Survey, financial losses due to network intrusion in the US alone exceeded $450 Million in 2002. But these figures only represent financial losses that survey respondents were willing and/or able to quantify and report. Computer Economics (www.computereconomics.com), for example, estimated that the world wide economic impact of the four viruses Code Red, Nimda, SirCam, and I LoveYou was $2.62, $1.15, $0.64, and $8.75 billion, respectively. They also identified the top five types of attack: Virus, Laptop Theft, Net Abuse, Denial of Service (DoS) and System Penetration (SP). Among them, the DoS and SP attacks show the greatest increase between 1997 and 2002. Given the increasing dependence of modern society on the use of the Internet, it is alarming to learn that probably less than 4% of network attacks were actually detected or reported.

Denial of Service is particularly interesting because of its huge impact on e-commerce systems or critical systems such as national defense systems. Despite these losses are incalculable. For e-commerce companies, DoS attacks reduce the trust and loyalty of customers and causes business loss. If the national Internet backbone is being attacked, DoS attacks may halt the Internet activities of an entire country.

Rule-based systems are most widely deployed in network intrusion detection products. They are easy to understand and use, but require human domain experts to find the rules and their generalization power depends on the expertise knowledge in the attacks. Machine learning and data mining techniques are possible solutions to this drawback, but this heavily depends, again, on the domain experts to tell what features are important to learn [1].

No matter what technique is used to deal with network intrusion detection, the features under study are the major problem for researchers. In this paper, we introduce a feature importance ranking methodology using stochastic radial basis function neural network sensitivity measure (RBFNN-SM).

In the next section, we discuss the DARPA data set. In Section 3, the proposed RBFNN-SM feature importance ranking will be introduced with experiments showing the accuracy of both full and reduced sets. The results are further analyzed in Section 4 and the last section concludes this paper.

2 DARPA Network Intrusion Detection Data Set

The DARPA network intrusion detection data set was prepared by MIT Lincoln Labs for the 1998 KDD Cup contest. The raw data came from the TCP dump data for a LAN environment that simulated a U. S. Air Force LAN for nine weeks. This raw binary data was processed to generate approximately five million connection records for the first seven weeks. The other two weeks' were processed to about two million connection records. Interesting facts are that the testing set contains some types of attacks that did not exist in the training set and it is not necessary that the probability density function of classes in the training and

0-7803-7865-2/03/$17.00 ©2003IEEE

testing sets be the same. This is to simulate the real world situation that novel, unseen attacks are common on the Internet. But, this paper does not seek to construct an optimal network intrusion detector to deal with unseen attacks. We leave that to future work. Our goal is to evaluate the features for DoS attack classification using RBFNN-SM, only the 10% training set of the full set is used and it was downloaded from the UCI database [7]. Furthermore, as 99% of records in the training set (also true to the testing set) are DoS attacks / Normal and DoS attack is a very serious problem on the Internet, we focus on DoS attack only in this paper. This means only the Normal and six classes of DoS attack are used in the experiments and analysis of this paper. The remaining 5285 records are removed from the data set. Furthermore, the data set was normalized to [0,1] for RBFNN while maintaining the original distribution shape using the following formula.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

There are in total 494,020 records, unevenly distributed across 23 classes. 98.93% of them fall into Normal and DoS, while 98.23% of total records fall into classes Normal, Smurf and Neptune. Statistically, the data set is dominated by Normal and DoS attacks, especially the three classes mentioned.

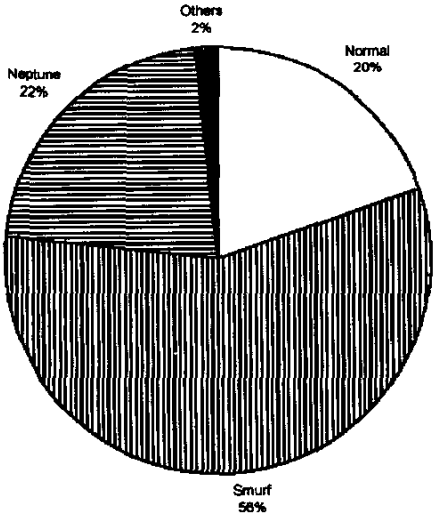

Figure 1. Class Distribution in the Full 10% Training Set

The 23 classes were categorized into 5 categories: Normal, DoS, Unauthorized Access from Remote Machine, Unauthorized Access to Local Super user and Probing.

Denial of Service	Smurf Neptune Back Land Teardrop Pod (Ping of Death)
Unauthorized Access from Remote Machine	Ftp_write Guess_passwd Imap Multihop Phf Spy Warezclient Warezmaster
Unauthorized Access to Local Superuser	Buffer_overflow Loadmodule Perl Rootkit
Probing	Ipsweep Nmap Portsweep Satan

Table 1. Categories of Attacks

There are in total 41 features, suggested by domain experts, in the data set. The one by one description of them will be presented together with the RBFNN-SM ranking in a later section. The features are generated in three ways:

1 Basic features of individual TCP connections (e.g. "Duration" and "Flag")
2 Content features within a connection suggested by domain knowledge (e.g. "Number of Files Accessed" and "Is It Guest Login")
3 Traffic features computed using a two-second time-window (e.g. "same service rate" and "SYN error rate")

The original data set downloaded contains non-numerical features: Protocol, Services and Flag. They are converted into numeric by assigning number to represent each of them, for example, 1 for TCP, 2 for UDP and 3 for ICMP for the "Protocol" feature.

3 Feature Importance Ranking Using RBFNN-SM

3.1 Why Do We Need Feature Selection?

In practice, the huge amount of data flowing on the Internet makes the real-time intrusion detection nearly impossible. Even though computing power is increasing exponentially, Internet traffic is still too big for real-time

computation. Feature selection can reduce the computation and model complexity. This makes it easier to understand and analyze by human and more practical to launch real-time intrusion detection system in large networks. Furthermore, the storage requirements of the data set and the computational power of generating indirect features, such as traffic signature and statistics, can be reduced by feature reduction.

Sensitivity analysis is a fundamental tool to analyze a neural network input-output relationship. Stochastic RBFNN-SM measures the output perturbation of an ensemble of radial basis function neural networks with respect to the input and weight perturbations (weight perturbations are set to zero in feature selection problems). Ng et al. extended this analysis tool to feature selection in [6]. Details of the stochastic RBFNN-SM may be found in [5, 6].

In this paper, the RBFNN-SM is used to find the correlation between input features and the output. A feature with high sensitivity means it is highly correlated to the output of RBFNN. Our aim in applying the feature ranking is to remove insignificant features for DoS detection in order to increase the generalization power of the classifier being implemented and reduce its complexity by reducing the number of parameters of the model.

3.2 Methodology of Computing RBFNN-SM

First of all, k-means clustering was performed to find the centers of a mixture of Gaussians that represents the data set, that is the DARPA 10% intrusion data set containing all Normal and six classes of DoS attacks. Ten clusters were assigned to each class with a predefined width of $\sqrt{0.1}$. The RBFNN-SM is computed using an ensemble of RBFNN for the same problem. Their architectures were all the same with different connection weights. Their connection weights were randomly selected from a uniform distribution between [-1, 1]. Two thousands were randomly selected.

Then the statistics of the data set, for example mean and variance of each input feature, together with the center matrix, widths, mean and variance of the weight vectors and statistics of input perturbations were fed into the stochastic RBFNN-SM formula to compute the sensitivity measure of features. For computing each feature sensitivity, this particular feature was perturbed randomly around 10% of original value 10 times for each sample in it, while keeping perturbations of all other features at zero. After calculating all the 41 features, one can rank the features in order of significance of the feature to the output of RBFNN.

Feature Name (Feature ID)	*RBFNN-SM*
Destination Host Count (32)	0.0030
Destination Host Service Count (33)	0.0026
Destination Host Same Service Rate (34)	0.0024
Same Service Rate (29)	0.0020
Destination Host Same Source Port Rate (36)	0.0020
Count of # Connections to Same Host Within 2 Seconds (23)	0.0018
Protocol Types (2)	0.0018
Service Count Within 2 Seconds (24)	0.0018
% of Connection Having SYN errors (SERROR_RATE) (25)	2.0023e-004
% of Service Having SYN errors (SRV_SERROR_RATE) (26)	1.9698e-004
Destination Host SYN error rate (38)	1.9416e-004
Destination Host Service SYN error rate (39)	1.9301e-004
Logged in (12)	1.5716e-004
REJ Error Rate (27)	2.5433e-005
Service REJ Error rate (28)	2.5395e-005
Destination Host Service REJ Error Rate (41)	2.4853e-005
Destination Host REJ Error Rate (40)	2.4347e-005
Service Types (3)	1.7797e-005
Service Different Hosts Rate (31)	4.4924e-006
Destination Host Different Service Rate (35)	1.0952e-006
Different Service Rate (30)	2.9508e-007
Wrong Fragments (8)	4.6210e-008
Destination Host Service Different Host Rate (37)	4.2531e-008
Flag (4)	4.1103e-008
Is Guest Login (22)	3.2026e-008
Number of Hot Indicators (10)	1.2323e-008
Duration (1)	2.3567e-009
Is Root Shell Obtained (14)	1.7718e-009
Is Connection from the Same Host / Port (LAND) (7)	6.9960e-010
Number of Data Bytes from Destination To Source (6)	6.4372e-010
Number of Shell Prompt (18)	4.7510e-010
Number of File Accessed (19)	3.2330e-010
Is SU ROOT Command Attempted (15)	2.3780e-010
Number of Files Created (17)	1.8656e-010
Number of Failed Login Attempted (11)	1.4884e-010
Number of Compromised Conditions (13)	6.4801e-011
Number of Root Accesses (16)	6.3324e-011
Number of Urgent Packets (9)	5.3177e-011
Number of Data Bytes from Source To Destination (5)	3.1548e-011
Is the Login Belongs to the Hot List (21)	2.8422e-014
Number of Outbound Commands in an FTP Session (20)	2.8422e-014

Table 2. Feature Ranking and Full names of the 41 Features

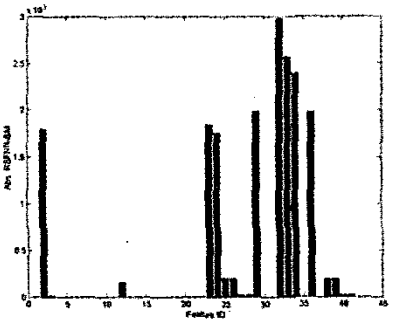

Figure 2. Sensitivity of Each Feature

3.3 RBFNN Classification Steps and Results

The data set was randomly separated into two halves for training and testing of the RBFNN. To prevent some small classes missing in either training or testing set, the 50% random selection was performed by class.

Two experiments were performed to test the reduced feature subset. Experimental results show that the reduced subset maintains the training and testing accuracy well after removing 33 features that is over 80% reduction in the data size.

In the first experiment, a binary classification task was performed using RBFNN to separate Normal records and DoS attack records. In the first round, all 41 features were used in the supervised training of the RBFNN and for the testing to validate the RBFNN. In the second round, only 13 high sensitivity features were used. The 8 most significant sensitive features were used in the last round. Experimental results were shown in the Table 3. The false alarm rate was computed by one minus the normal classification accuracy, and the false positive rate was defined as one minus the DoS Attack classification accuracy.

It is promising that the classification accuracy can be maintained on the same high level with only 8 features while incurring less computation complexity. The false alarm rates for the 8 features subset (41 features full set) are 0.18% (0.01%) and 0.27% (0.03%) in training and testing, respectively. While the false positive rates are 0.93% (0.70%) and 0.94% (0.71%). This shows that the reduced data set gives comparable results to the original full set, especially in the false positive rate, with approximately 3 times less computational effort (reduced from 23 seconds training to 9 seconds).

	Normal	DoS Attacks	Time Used
41 Features Training	99.99%	99.30%	23 seconds
41 Features Testing	99.97%	99.29%	20 seconds
13 Features Training	99.82%	99.28%	11 seconds
13 Features Testing	99.78%	99.26%	7 seconds
8 Features Training	99.82%	99.07%	9 seconds
8 Features Testing	99.73%	99.06%	6 seconds
# Samples In Training	*48639*	*195731*	
# Samples In Testing	*48638*	*195727*	

Table 3. RBFNN Classification Accuracy for Normal / DoS Attack Detection Using Reduced Feature Subset

In the second experiment, we tried to classify not only the DoS attacks, but also what type of DoS attack was launching by the connection or packet. The steps are the same as the first experiment, except that this time the RBFNN outputted seven classes.

Be aware that there are 4 extremely small classes; totally contribute only 0.7% of the records in the DoS category. RBFNN usually cannot classify well for totally unbalanced data set. This may be solved by some other techniques, such as training more on not well-trained portion [3]. But this is not the focus of this paper, so we leave them to future works. The results are shown in the Table 4. One may notice that the Neptune classification is improved with fewer features, while the Normal class is a little bit worse than the full set. The false alarm rate raised from 2.29% to 3.20%, which is very insignificant change, while more than 80% of features was removed. The computation time for RBFNN testing decreased from 70 seconds to 20 seconds for about 245,000 records. The classification accuracy for Neptune rose from 94.93% to 99.27% in 8-feature subset, while the Smurf type remained unchanged generally. So, the training and testing accuracies were overall remaining the same high accuracy for those dominated classes.

When comparing the results of both experiments, the training and testing accuracies dropped very insignificantly after reducing the 41 features to 8 features. The experiments proved the feature subset selected by RBFNN-SM ranking is good enough to classify the Normal and DoS attack packets and connections.

	41 Features	13 Features	8 Features	# Samples
Training				
Time Used	95 seconds	55 seconds	46 seconds	
> Normal	97.71%	97.25%	96.80%	48639
> Smurf	99.99%	99.99%	99.99%	140395
> Back	10.07%	0.00%	0.00%	1102
> PoD	0.00%	0.00%	0.00%	132
> Neptune	95.09%	98.10%	99.29%	53601
> Teardrop	34.29%	23.88%	20.82%	490
> Land	72.73%	72.73%	0.00%	11
Testing				
Time Used	70 seconds	28 seconds	20 seconds	
> Normal	97.61%	97.32%	96.87%	48638
> Smurf	100.00%	99.99%	99.99%	140395
> Back	8.99%	0.00%	0.00%	1101
> PoD	0.00%	0.00%	0.0152%	132
> Neptune	94.93%	98.04%	99.27%	53600
> Teardrop	32.72%	24.74%	22.49%	489
> Land	50.00%	30.00%	0.00%	10

Table 4. RBFNN Classification Accuracy and Class Distribution

4 Discussions

4.1 Discussion on the Features Selected

4.1.1 Selected Features and Smurf Attack

The experimental results show that the 8 features selected is good enough for classifying Normal and DoS attacks. The 8 most sensitive features include the "Protocol Type", "Service Count" (number of connection requesting the same service within two seconds), "Number of Connection Connecting to the Same Host", "Same Service Rate, Destination Host Count", "Destination Host Same Service Rate", "Destination Service Count" and "Destination Host Same Source Port Rate". They are significant in classifying Smurf attack. Smurf attack is to create ICMP subnet directed broadcast to announce the victim's IP address and let all hosts on the LAN to reply to the victim. So many packets with same service request are sent to the same host simultaneously and all of the packets are using ICMP.

4.1.2 Selected Features and Neptune Attack

Neptune was first discovered in 1996. It launches a SYN flooding attack against a victim by sending session establishment packets using forged source IP addresses. When the rate of the SYN packets is high enough, the victim's resource is used up to wait for the session to be confirmed [2]. Out of the 8 most sensitive features mentioned in section 4-A-1, 5 of them, namely, "Destination Host Same Service Rate", "Same Service Rate", "Destination Host Same Source Port Rate", "Service Count Within 2 Seconds" and "Number of Connections to Same Host Within 2 Seconds", are all symptoms of SYN flooding attacks. These features are related to the nature of SYN flooding attack that it sends the same request to bomb-up the victim. Moreover, 4 out of the 5 extra features among the 13 most sensitive ones are also related to SYN error, which is highly correlated to SYN flooding type DoS attacks.

4.1.3 Selected Features and General DoS Attack

The "Logged In" feature in the 13 feature subset seems not closely related to classifying DoS from Normal. But, remember that the original data set has 23 classes, those non-DoS attacks are mainly unauthorized remote user hacking and superuser privilege gain. They must be logged in, so this feature helps to distinguish the DoS from the other types of attack.

4.1.4 Further Discussion on the Input Features

One may notice that the 28 most insensitive features are not correlated to the DoS attack classification. For examples, the features "Is It Guess Login", "Does It Attempt to Use Superuser Commands" and "Number of File Created", are more likely to be correlated to unauthorized access. So, this is insignificant to DoS detection.

As shown in the previous section, the Smurf, Neptune and Normal classes dominate the data set. The sensitivity measure computed for the data set is actually the sensitivity of input features with respect to Normal / DoS attacks (especially Smurf and Neptune). This is because the derivation of the RBFNN-SM is bias to dominated classes if the data set is totally unbalanced.

Mukkamala et al [4] selected 19 and 11 features that they claim are important to DoS attack detection in two experiments by using a brute force method. Their selected features are: "Duration", "Number of Data Bytes from Source to Destination", "Number of Data Bytes From Destination to Source", and features 23, 24, 25, 26, 32, 36,

38 and 39. The last 8 are also important features selected by our sensitivity-based method and they had already been discussed. But the other features seem to be not so much related to DoS attacks, but rather to Unauthorized Access from Remote Machine, such as FTP_Write attack. Unfortunately, the authors did not explain how they selected the 5092 training samples and what pre-processing had been done to the data set, so we cannot repeat their experiments for comparison. In addition, they did not present any discussion on the relationship between the selected features and DoS attacks.

We may conclude that the RBFNN-SM helps to identify features that are more important to DoS attack classification in the DARPA 10% intrusion detection data set. This is supported by both experimental results and by domain experts' knowledge. The benefits of this reduction are the simpler model for DoS detection, faster training and testing, reduced storage requirements (less than 20% data are needed in terms of byte) and better generalization capability.

4.2 Discussion on the Small Classes Accuracy

The classification results for binary classes are better than those for 7 classes. This is reasonable because binary classifier has less complexity with the same number of features when compared to 7 classes. Especially for those small classes, the extreme case is the Land class, which contains only 21 records in total, compared to the 494,020 records in the full set. So, the classification results fluctuate widely. If users want a higher accuracy in all classes, one simple idea is to divide them into 7 RBFNNs, with each one dealing with only 1 class, instead of using a single RBFNN classifying 7 classes. Another suggestion is to feed the feature subset selected based on the feature ranking using RBFNN-SM into other classifiers different from the RBFNN. What we attempt to do in this paper is to use a single tool for both feature selection and classification.

5 Conclusion

In this paper, we have applied the stochastic RBFNN-SM to rank the feature importance of DARPA 10% intrusion detection data set for DoS attack classification. Experiments show that our method reduces 33 features out of 41 (80.49%). The remaining 8 features maintain comparable training and testing accuracies to the original full set. In addition, some observations are made on the high correlation between the features selected and the DoS attacks, especially Smurf and Neptune.

The proposed RBFNN-SM feature ranking method could potentially be used as an effective tool for general network intrusion detection problems, and the resulting feature subset may be compared with those obtained from expert domain knowledge. This may help to find better problem description and provide evidence for feature selection for modeling network intrusion.

Acknowledgements

This work was partially supported by Hong Kong CERG grants #B-Q571 and #POLYU5080/02E.

References

[1] D. Barbará and S. Jajodia, "Applications of Data Mining in Computer Security", Kluwer Academic Publishers, 2002.

[2] R. K. C. Chang, "Defending Against Flooding-Based, Distributed Denial-of-Service Attacks: A Tutorial," IEEE Communications Magazine, vol. 40, no. 10, pages 42 – 51, 2002.

[3] S. Haykin, "Neural Networks: A Comprehensive Foundation", Prentice Hall, 1999.

[4] S. Mukkamala and A. H. Sung, "Detecting Denial of Service Attacks Using Support Vector Machines", IEEE Proc. Int. Conference on Fuzzy Systems, pages 1231 – 1236, 2003.

[5] W. W. Y. Ng, D. S. Yeung, Q. Ran and E. C. C. Tsang, "Statistical Output Sensitivity to Input and Weight Perturbations of Radial Basis Function Neural Networks", IEEE Proc. of Int. Conferences on System, Man and Cybernetics, pages 503 – 508, 2002.

[6] W. W. Y. Ng and D. S. Yeung, "Input Dimensionality Reduction for Radial Basis Neural Network Classification Problems Using Sensitivity Measure", Proc. on Int. Conference on Machine Learning and Cybernetics, Beijing, pages 2214 - 2219, 2002.

[7] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html